EMPIRICAL DISTRIBUTIONS OF STOCK RETURNS AND APPLICATIONS IN VALUE AT RISK

by

Difei Zhu
Bachelor of Economics, South China Normal University, 2013

and

Jiaxiu Sun
Bachelor of Economics, Tianjin Normal University, 2013

PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN FINANCE

In the Master of Science in Finance Program
of the
Faculty
of
Business Administration

SIMON FRASER UNIVERSITY

Fall 2014

Approval

Name: **Difei Zhu and Jiaxiu Sun**

Degree: **Master of Science in Finance**

Title of Project: **Empirical distributions of stock returns and applications in Value at Risk**

Supervisory Committee:

____

Andrey Pavlov
Senior Supervisor
Professor, Faculty of Business Administration

____

Kangmin Lee
Second Reader
Sessional Lecturer, Faculty of Business Administration

Date Approved: ___________________________________

Abstract

This paper examines the best distribution for stock returns. Normal distribution is the basic assumption we assume when understanding and analyzing different kinds of models. However, this does not always work when describing stock returns distribution because they always have fat tails that cannot be explained. We used several databases and distributions to find out the most suitable distribution for stock returns by examining them through A-D test and Value at Risk application. The result is that Generalized Pareto Distribution is the most suitable one statistically.

Key words: Return distribution; VaR; Fat tails; A-D test; Generalized Pareto distribution

Acknowledgements

We would like to express our sincere gratitude to our supervisor, Dr. Andrey Palvov, who offered us valuable feedbacks, encouragement and patience to support us throughout this project. He also provided us with stimulating comments and suggestions. With his rich research experience, he guided us the research direction we should go forward.

Also, we would like to give our special thanks to Phil Goddard in assisting us about programing language and to Khangmin Lee for the help to our project.

Table of Contents

1. Introduction

The empirical distributions for stock returns are drawing more attention because more papers and documents have been devoted. The shape of the tail of the returns distribution is having more implications rather than ignored by some economists in the past. In this paper, we are trying to find out the most suitable distribution for the stocks returns.

To reach this goal, we first gather some raw data for analyzing. The databases applied here are comparatively representative for the current economy like the Dow Jones Industrial Average Index, daily and minute-by-minute S&P 500 Index and the MSCI Word Index. Secondly, we graph the returns distribution curve for each database and compare each real distribution with the normal distribution. As all the return distribution curve are significantly different from the normal distribution curve, we used some other distributions like Weibull distribution to fit the fat tails and leptokurtosis

Finally by applying the Anderson-Darling test we find that the Generalized Pareto Distribution and Log-Weibull distribution are much better than the other distributions we tested. Through some data analysis, we suggest that the Generalized Pareto Distribution (also called Paretian Distribution or stable distribution) is superior to the Log-Weibull distribution by using Value at Risk in statistical analysis.

2. Literature Review

In most financial models, when it comes to the question, how to use the historical stock prices to make meaningful predictions to the future price, the answers are separated into two primary streams, one is the various chartist theories stream and the other one is the random walk theory stream. For the chartist theories, they all assume that the past behavior of the stock will heavily impact the future behavior, that is, the historical price “patterns” will repeat itself to recur in the future trend. Thus once we find out these “patterns” through analyzing the price charts, we can use them to forecast the future behavior of the stock prices.

However, on the opposite side, the random walks point out that the past cannot be used to predict in any meaningful way because the price changes are independent and identical distributed random walks.

2.1 Random Walk stream representatives

Bachelier is the first one to complete the development of a random walks in security prices. His model was derived by Osborne 50 years later to recommend that under several assumptions such like price changes are IID, the distribution of price changes has finite variance and the price changes will be sums of many independent variables. The price changes will each have normal distribution. This model is then called Bachelier-Osborne model.

Similarly, Moore and Kendall also provided empirical evidence to support the Gaussian hypothesis. Moore graphed the weekly first difference of log price of 8 NYSE common stocks and Kendall observed several British common stocks. They both made great effort to support the hypothesis of approximate normality. However they dropped some extreme tails from their subsequent statistical tests and ignored the fact that most of the distributions of price changes are leptokurtic which means too many values near the mean and too many values out in the extreme tails.

2.2 Other distributions representatives

Even though the research result is based on so many assumptions that may be not coherent with the real economy situation, the Gaussian hypothesis was not questioned until early 1960s. In 1963 Benoit Mandelbrot proposed that the past academic research has neglected the leptokurtosis on purpose. If the outliers are numerous, neglecting them takes away much too significance from the tests carried out from the data. Mandelbrot put forward that the Stable Paretian Distribution is more appropriate for returns distribution.

Two years later, in 1965, Eugene F Fama testified Mandelbrot’s statement and summarized that the first difference of stock prices of stock prices seem to follow stable Paretian Distribution with characteristic exponent less than 2. By testing the empirical validity of the random walk models, he also draw the conclusion that the chart reading through the past will not contribute any valuable information to the stock market investor.

Later on, many distributions were tested. In 1994, Stuart and Ord identified the Student t distribution with about 4.5 degrees of freedom as the best fit to the observed daily log-returns of the S&P 500. In 2006, Kevin Fergusson and Echhard Platen confirmed this conclusion. They tested the student t distribution by using the daily log-returns of the Word Stock Index over the period from 1970 to 2004. In 1996, Markowitz and Usman analyzed 20 years log-returns of daily S&P 500 covering the period from 1963 to 1983 in a Baysian framework. In 1997, Stefan and Svetlozar discussed Weibull distribution turned out to be more appropriate distribution in fitting fat tails. In 2005, Maleverge, Pisarenko and Sornette developed a battery of new non-parametric and parametric tests to characterize the distributions to propose the log-Weibull model is an appropriate approximation of the return distributions.

Even though a huge number of distributions were tested during the past hundreds of years, a distribution that generally fits log-returns of stock indices has not been widely agreed on so far.

3. Data and Methodology

To analyze the stock returns distribution, we chose Dow Jones Industrial Average Index, daily and minute-by-minute S&P 500 Index and MSCI World Index as the representatives for the stock market.

3.1 Data and Basic Analysis

Dow Jones is short for Dow Jones Industrial Average Index which is a price weighted average of 30 significant stocks traded in New York Stock Exchange or NASDAQ. In this paper, we use the Dow Jones data over the time period from May 27, 1896 to May 31, 2000. This time period is always analyzed by some related econometricians, such as Longin(1996) and Malevergne, Pisarenko, Sornette (2004).

S&P 500 Index is short for Standard & Poor's 500 which is based on the market capitalization of the first 500 large companies that have common stocks traded in NYSE or NASDAQ. Here for the daily S&P 500, we use data from Jan 2 in 1926 to Nov 13 in 2014. For the minute-by-minute one, we use data form May 2 in 2014 to Nov 13 in 2014.

MSCI world index is an index of 1621 word's stocks and usually used as a common benchmark for world or global stock funds. We use the data from July 3 in 1962 to Nov 13 in 2014. The table 1 below is the outcome of our data calculation.

Table 3.1 Data and basic results for the 4 indices

	Time Period	**Mean (*1000)**	**Standard deviation (*100)**	**Skewness**	**Kurtosis**	**Jarque-Bera test p-value (*1000)**	**Jarque-Bera test statistic (/100,000)**
Dow Jones	May 27, 1896 - May 31, 2000	0.21	1.1	-0.63	22.34	1.0	4.1
S&P 500	Jan 02, 1926 – Nov 13, 2014	0.32	1.0	-0.64	24.07	1.0	2.4
MSCI World Index	July 3, 1962 – Nov 13, 2014	0.42	1.1	-0.03	19.70	1.0	2.7
S&P 500 (Minute-by-Minute)	May 2, 2014 – Nov 13, 2014	0.14	0.034	0.29	48.55	1.0	48

As it showed in table 3.1, we first calculated some basic elements of probability distribution such as the mean, standard deviation, Skewness and Kurtosis for each of the index and even made the Jarque-Bera test for these outcomes.

3.1.1 Mean and Standard deviation

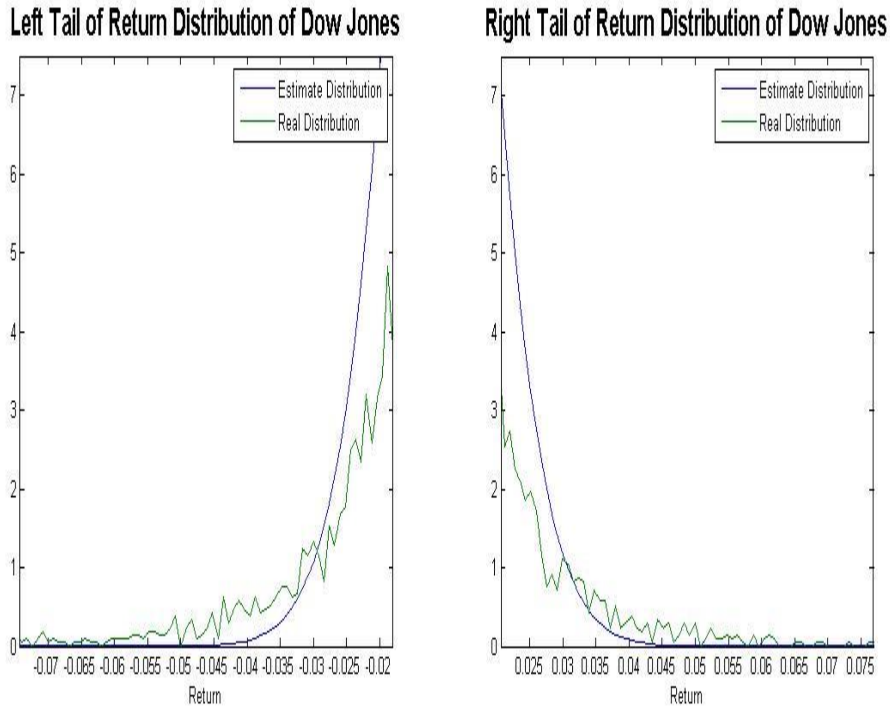
Mean is calculated to reflect the central value for a series of data. Standard deviation measures the amount of variation or dispersion from the mean, and the lower this number is, the more concentrating the data points are. From table3.1 we can state that the standard deviation of the first 3 indices are very similar to each other. The minute by minute S&P 500 has a much smaller standard deviation.

3.1.2 Skewness

Skewness measures the asymmetry of all the data. In general, a positive Skew means the tail on the right side is longer and fatter than the left side and a negative Skew indicates the a longer and fatter tail on the left side than the right side. The longer and fatter the tail of a distribution is, the more extreme values it contains. For the indices we use, the Dow Jones and daily S&P 500 have the similar skew, -0.63, -0.64 respectively, the MSCI also have a negative index of -0.03 but the 1-minute S&P 500 has a positive skew of 0.29 which means the right side tail is fatter, but not including the same amount of extreme values as the Dow Jones and daily S&P contain.

Figure 3.1 below are the left tail and right tail of the returns distribution of Dow Jones. All the indices we use have the similar tails as these two, for simplicity we only use them to show the result. The fatter left tail shows the left side is more incoherent from the normal distribution than the right side.

Figure3. 1 Left tail and right tail of the returns distribution the Dow Jones Index



3.1.3 Kurtosis

Kurtosis describes the shape of a probability distribution. If the value is positive, the shape is steeper compared with the appearance of normal distribution. If the value is negative, the shape is flatter. The higher the absolute value is, the more different it is from the normal distribution. For our data, all the indices have the positive Kurtosis, especially for the 1-minute S&P which has the Kurtosis as high as 48.55.

3.1.4 Jarque-Bera test

Jarque-Bera test is to find out whether the sample data have the Skewness and Kurtosis matching the normal distribution. The null hypothesis of this test is that the data is following a normal distribution. When the probability is higher than 0.05, we cannot reject the null hypothesis. When the probability is smaller than 0.05, we reject the null hypothesis. We can clearly find the J-B test probabilities of all the 4 indices are all much smaller than 0.05. None of the 4 returns distributions follows a normal distribution.

3.2 Data analysis

Besides the results and outcomes, we even split the stock returns into positive return and negative return for each of the index. These returns are the log-returns of the original data. For the negative returns, we get the absolute values and then use the same method to analyze.

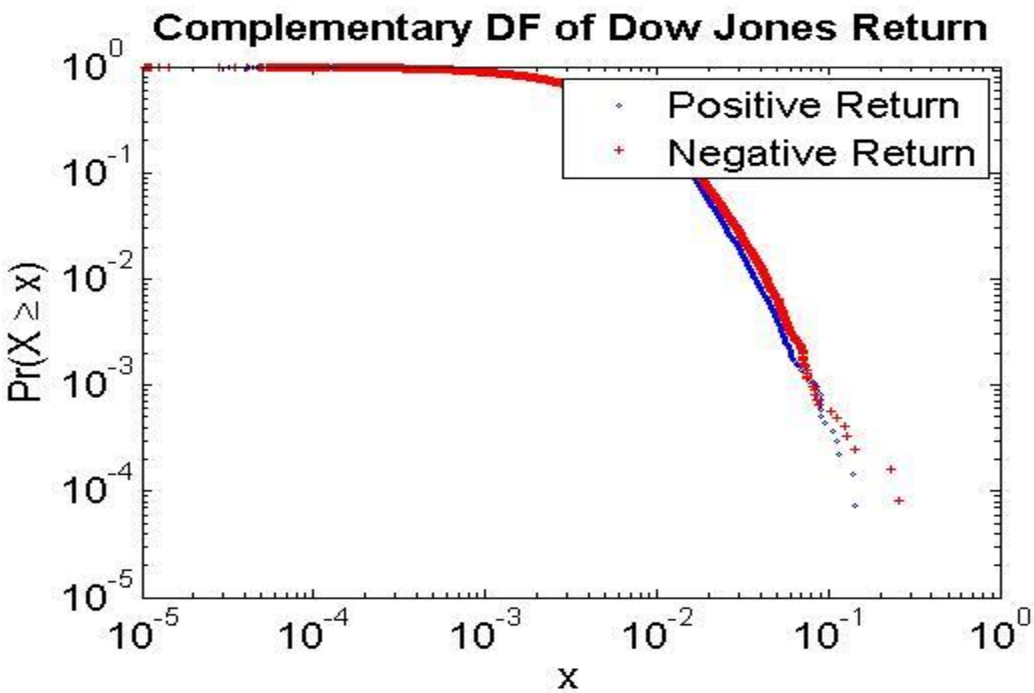
In table 3.2, the threshold probability determines threshold return. For example, in the Dow Jones index table, when standard threshold probability is 0.2, corresponding threshold return for positive return is 0.1959% and that for negative return is 0.1823%. The used samples contains those positive returns that are not smaller than the threshold return. The number of data used which is10860 provides the size of the samples beyond the positive threshold return 0.1959% and 9804 means the number of samples beyond 0.1823% for the negative returns.

. We also graphed the complementary cumulative distribution functions for Dow Jones returns which is titled figure 3.2. In the figure, the horizontal axis is the return, the vertical axis is the probability that the return is higher than the corresponding threshold return, namely $Pr(X \geq x)$ where X denotes log-returns and x represents threshold returns. We can observe the coherent result as the table 2 shows. The same analysis to the other 3 indices see Appendix B

Table 3.2 Probability levels, corresponding lower thresholds and size of sub-sample beyond the thresholds of Dow Jones Industrial Index from May 27, 1896 to May 31, 2000

Dow Jones				
	Positive Return		Negative Return	
Threshold (probability)	Number of Data used	Threshold (Min. Return *100)	Number of Data used	Threshold (Min. Return *100)
0	13575	0.0026	12255	0.0023
0.1	12218	0.1033	11030	0.0910
0.2	10860	0.1959	9804	0.1823
0.3	9503	0.3009	8579	0.2763
0.4	8145	0.4108	7353	0.3868
0.5	6788	0.5315	6128	0.5142
0.6	5430	0.6734	4902	0.6771
0.7	4073	0.8641	3677	0.8716
0.8	2715	1.1045	2451	1.1660
0.9	1358	1.5758	1226	1.6918
0.925	1018	1.7716	919	1.9525
0.95	679	2.0796	613	2.3742
0.96	543	2.2789	490	2.5964
0.97	407	2.5515	368	2.9613
0.98	272	3.0287	245	3.4419
0.99	136	3.7742	123	4.3599
0.9925	102	4.1936	92	4.8120
0.995	68	4.7949	61	5.3698

Figure 3.2 Complementary Cumulative Distribution Function of the Dow Jones Index



3.3 Statistic Methodologies Introduction

This part will focus on describing some famous distributions used by researchers in fitting fat tails. Five models are tested, including generalized Pareto distribution, Weibull distribution, exponential distribution, power law distribution and log-Weibull distribution.

3.3.1 Generalized Pareto Distribution

Pareto Distribution is proposed by Vilfredo Pareto, a famous Italian engineer, socialist and economist. He found out the fact that incomes of individuals in Italian follow a distribution known as 80-20 rule. Nowadays, it is widely used in describing incomes of individuals, size of residential area, amount of oil reserves in oilfield and empirical returns of stocks and commodity.

Cumulative Distribution Function

$$F_X(x) = \begin{cases} 1-(1+k*\frac{x-x_{min}}{\sigma})^{-\frac{1}{k}}, & x \geq x_{min} \\ 0, & x < x_{min} \end{cases}$$

Probability Density Function

$$f_X(x) = \begin{cases} \frac{1}{\sigma} * (1 + k * \frac{x - x_{min}}{\sigma})^{-1-\frac{1}{k}}, & x \geq x_{min} \\ 0 \quad , & x < x_{min} \end{cases}$$

K is shape parameter. σ represents scale parameter. If sharp parameter, K, and $x_{min} = 0$, then generalized Pareto Distribution converges to exponential distribution. In addition, if $x_{min} = \sigma/k$, generalized Pareto distribution converges to Pareto distribution.

3.3.2 Weibull Distribution

Weibull distribution is named after Waloddi Weibull, a Swedish engineer and mathematician. It is used by Stefan and Svetlozar (1997) in fitting returns distribution.

Cumulative Distribution Function

$$F_X(x) = \begin{cases} 1 - \exp[-\left(\frac{x}{d}\right)^c + \left(\frac{x_{min}}{d}\right)^c], & \mathrm{x} \geq x_{min} \\ 0 \quad , & \mathrm{x} < x_{min} \end{cases}$$

Probability Density Function

$$f_X(x) = \begin{cases} \frac{c}{d^c} * x^{c-1} * exp[-(\frac{x}{d})^c + (\frac{x_{min}}{d})^c], & x \geq x_{min} \\ 0 \quad , & x < x_{min} \end{cases}$$

c reflects shape. d is scale parameter. If shape parameter c equals 1, Weibull distribution is the same as exponential distribution.

3.3.3 Exponential Distribution

Exponential distribution has the important property of memory less. It is used in explaining time interval of independent random events such as income phone calls.

Cumulative Distribution Function

$$F_X(x) = \begin{cases} 1 - \exp(-\frac{x}{d} + \frac{x_{min}}{d}), & \mathrm{x} \geq x_{min} \\ 0 \quad , & \mathrm{x} < x_{min} \end{cases}$$

Probability Density Function

$$f_X(x) = \begin{cases} \frac{1}{d} * \exp(-\frac{x}{d} + \frac{x_{min}}{d}), & x \geq x_{min} \\ 0 \quad , & x < x_{min} \end{cases}$$

1/d is the rate which is always expressed by λ. Also, cumulative distribution and probability density function of exponential distribution are modified by Y. Malevergne, V. Pisarenko and D. Sornette(2005) with a non-zero threshold.

3.3.4 Log-Weibull Distribution

Y. Malevergne, V. Pisarenko and D. Sornette(2005) suggested to use two-parameter log-Weibull distribution to fit the heavy tails of empirical returns. This distribution interpolates between Weibull distribution and generalized Pareto distribution.

Cumulative Distribution Function

$$F_X(x) = \begin{cases} 1 - \exp[-b * ln\left(\frac{x}{x_{min}}\right)^c], & \mathrm{x} \geq x_{min} \\ 0 & , \mathrm{x} < x_{min} \end{cases}$$

Probability Density Function

$$f_X(x) = \begin{cases} \frac{b*c}{x} * ln\left(\frac{x}{x_{min}}\right)^{c-1} * \exp[-b * ln\left(\frac{x}{x_{min}}\right)^c], & x \geq x_{min} \\ 0 & , x < x_{min} \end{cases}$$

3.3.5 Power Law Distribution

Power Law Distribution is widely used in physics, biology, social science and finance. Zipf's law and Pareto Distribution are two representative of the Power Law Distribution.

Cumulative Distribution Function

$$F_X(x) = \begin{cases} 1 - (\frac{x}{x_{min}})^{-\partial+1}, & x \geq x_{min} \\ 0 & , x < x_{min} \end{cases}$$

Probability Density Function

$$f_X(x) = \begin{cases} \frac{\partial - 1}{x_{min}} * (\frac{x}{x_{min}})^{-\partial}, & x \geq x_{min} \\ 0 & , x < x_{min} \end{cases}$$

∂ is known as scaling parameter.

3.4 Parameters estimating and Distribution tests

This part is to estimate the parameters. The distributions are tested to show whether they are really suitable or not.

3.4.1 Parameters Estimating

The method used in estimating parameters of different distribution function is maximum likelihood method which is selecting the set of parameters value to maximize likelihood function.

If probability density function is $f(x_i, \theta)$, where θ is estimated parameter and $X_1, X_2, X_3 \ldots X_n$ are samples, then likelihood function is $L(\theta) = \prod_{i=1}^{n} f(x_i; \theta)$. According to Reuy S Tsay(2002), log function is monotone so that the maximum likelihood estimates can be obtained through maximizing the log likelihood function instead of likelihood function, that is, $\ln L(\theta) = \prod_{i=1}^{n} \ln[f(x_i; \theta)]$. In order to estimate θ, take $\frac{d}{d\theta} \ln L(\theta) = 0$. Here, the value of θ that maximizes log likelihood function is maximum likelihood estimated parameter.

3.4.1.1 Generalized Pareto Distribution

Probability Density Function

$$f_X(x) = \begin{cases} \frac{1}{\sigma} * (1 + k * \frac{x - x_{min}}{\sigma})^{-1-\frac{1}{k}}, & x \geq x_{min} \\ 0 \quad , & x < x_{min} \end{cases}$$

Therefore, $\ln L(\sigma, k) = \sum_{i=1}^{n} \ln[f(x_i; \sigma, k)]$. Then find the set of $(\hat{\sigma}, \hat{k})$ that maximizes $\ln L(\sigma, k)$.

3.4.1.2 Weibull Distribution

Probability Density Function

$$f_X(x) = \begin{cases} \frac{c}{d^c} * x^{c-1} * exp[-(\frac{x}{d})^c + (\frac{x_{min}}{d})^c], & x \geq x_{min} \\ 0 \quad , & x < x_{min} \end{cases}$$

The maximum of log likelihood function is $\ln L(\theta) = \sum_{i=1}^{n} [f(x_i; c, d)]$. Thus, the estimated parameters $(\hat{b}, \hat{d})$ are the solutions of

$$\begin{cases} \frac{1}{c} = \frac{\frac{1}{n} * \sum_1^n (\frac{X_i}{x_{min}})^c * \ln(\frac{X_i}{x_{min}})}{\frac{1}{n} * \sum_1^n \left(\frac{X_i}{x_{min}}\right)^c - 1} - \frac{1}{n} * \sum_1^n \ln(\frac{X_i}{x_{min}}) \\ d^c = \frac{x_{min}{}^c}{n} \sum_1^n (\frac{X_i}{x_{min}})^c - 1 \end{cases}$$

3.4.1.3 Exponential Distribution

Probability Density Function

$$f_X(x) = \begin{cases} \frac{1}{d} * \exp(-\frac{x}{d} + \frac{x_{min}}{d}), & x \geq x_{min} \\ 0 & , \ x < x_{min} \end{cases}$$

The maximum of log likelihood function is $\ln \mathrm{L}(\theta) = \sum_{i=1}^{n} [\mathrm{f}(\mathrm{x_i}; \mathrm{d})]$. Therefore, $\hat{d}$ estimated by maximum likelihood method is equivalent to the value of $\frac{1}{n} * \sum_1^n X_i - x_{min}$.

3.4.1.4 Log-Weibull Distribution

Probability Density Function

$$f_X(x) = \begin{cases} \frac{b*c}{x} * ln\left(\frac{x}{x_{min}}\right)^{c-1} * \exp[-b * ln\left(\frac{x}{x_{min}}\right)^c], & x \geq x_{min} \\ 0 & , \ x < x_{min} \end{cases}$$

In order to estimate parameter $(\hat{b}, \hat{c})$ by maximum likelihood method, the set of parameters value maximize log likelihood function, namely $\sum_{i=1}^{n} [\mathrm{f}(\mathrm{x_i}; \mathrm{b}, \mathrm{c})]$. The solution of $\max \prod_{\mathrm{i}=1}^{\mathrm{n}} \ln[\mathrm{f}(\mathrm{x_i}; \mathrm{b}, \mathrm{c})]$ is

$$\begin{cases} b^{-1} = \frac{1}{n} * \sum_{i=1}^{n} (\ln \frac{X_i}{x_{min}})^c \\ \frac{1}{c} = \frac{\sum_{i=1}^{n} (\ln \frac{X_i}{x_{min}})^c * \ln(\ln \frac{X_i}{x_{min}})}{\sum_{i=1}^{n} (\ln \frac{X_i}{x_{min}})^c} - \frac{1}{n} * \sum_{i=1}^{n} ln(\ln \frac{X_i}{x_{min}}) \end{cases}$$

3.4.1.5 Power Law Distribution

Probability Density Function $f_X(x) = \begin{cases} \frac{\partial-1}{x_{min}} * (\frac{x}{x_{min}})^{-\partial}, & x \geq x_{min} \\ 0 & , x < x_{min} \end{cases}$

In order to estimate $\hat{\partial}$, maximize log likelihood function. $\mathrm{Ln\,L}(\partial) = \sum_{i=1}^{n}[\mathrm{f(x_i};\partial)] = \mathrm{n} * \ln(\partial-1) - n * \ln x_{min} - \partial * \sum_{i=1}^{n} \ln \frac{x_i}{x_{min}}$. Value of $\hat{\partial}$ equals to $1 + \frac{n}{\sum_{i=1}^{n} \ln \frac{x_i}{x_{min}}}$

3.4.2 Distributions tests

In order to test whether these distributions fit log-returns, Anderson-Darling test is applied here. Anderson-Darling test focuses on testing whether sample data comes from hypothesized distribution.

The null hypothesis of Anderson-Darling test is that sample data originates from a hypothesized distribution, such as normal distribution. Anderson-Daring test statistic, according to Anderson and Daring (1952), follows a formula that $\mathrm{ADS} = \mathrm{n} * \int \frac{[F_n(x)-F(x)]^2}{F(x)*[1-F(x)]} dF(x)$, where $F(x)$ represents theoretical cumulative distribution function, $F_n(x)$ represents empirical cumulative distribution function and n denotes numbers of sample. So Anderson-Darling distance follow this formula that $\mathrm{ADS} = -\mathrm{n} - \sum_{i=1}^{n} \frac{2i-1}{n} * \{\ln[F(X_i)] + \ln[1 - F(X_{n+1-i})]\}$, where $X_1 < X_2 < X_3 < \cdots < X_n$ are the ordered sample data and n denotes number of sample.

In order to determine whether null hypothesis should be accepted or rejected, p-value based on Anderson-Darling distance is compared with confidence interval which is 5% in most testing cases. If p-Value is smaller than 5%, null hypothesis is rejected with 95% confidence interval. If p-Value is larger than 5%, null hypothesis cannot be rejected with 95% confidence interval, that means the sample data comes from a hypothesized distribution.

4. Empirical Results

At the beginning of the paper, we set 18 possible values for the standard threshold value. Here in the table 4.1, we summarized the threshold probabilities to two ranges, that is, the range of threshold $u_1 - u_9$ and $u_{10} - u_{18}$. The outcomes of tests are expressed as A/9, where A is frequency of acceptances of null hypothesis and 9 is the number of Anderson-Darling testing in each range.

Table 4.1 Anderson-Darling Test outcomes for the 5 distributions

	Distribution				
Threshold return	Pareto Distribution	Weibull Distribution	Exponential distribution	Log-Weibull distribution	Power-Law distribution
Dow Jones Industrial Average Index Positive Daily Return from May 27, 1896 - May 31, 2000					
$u_1 - u_9$	2/9	0/9	0/9	2/9	0/9
$u_{10} - u_{18}$	9/9	9/9	3/9	9/9	9/9
Dow Jones Industrial Average Index Negative Daily Return from May 27, 1896 - May 31, 2000					
$u_1 - u_9$	3/9	0/9	0/9	2/9	0/9
$u_{10} - u_{18}$	9/9	9/9	5/9	9/9	7/9
Standard and Poor 500 Index Positive Daily Return from Jan 02 1926 - Dec 31 2013					
$u_1 - u_9$	5/9	0/9	0/9	2/9	0/9
$u_{10} - u_{18}$	9/9	9/9	3/9	9/9	7/9
Standard and Poor 500 Index Negative Daily Return from Jan 02 1926 - Dec 31 2013					
$u_1 - u_9$	6/9	0/9	0/9	1/9	0/9
$u_{10} - u_{18}$	9/9	9/9	4/9	9/9	7/9
MSCI World Index Positive Daily Return from July 3 1962 - Dec 31 2013					
$u_1 - u_9$	4/9	0/9	0/9	1/9	0/9
$u_{10} - u_{18}$	9/9	9/9	4/9	9/9	9/9
MSCI World Index Negative Daily Return from July 3 1962 - Dec 31 2013					
$u_1 - u_9$	5/9	0/9	0/9	2/9	0/9
$u_{10} - u_{18}$	9/9	9/9	5/9	9/9	9/9
Standard and Poor 500 Index Positive Minute Return from May 2 2014 – Nov 13 2014					
$u_1 - u_9$	4/9	0/9	0/9	0/9	0/9
$u_{10} - u_{18}$	9/9	4/9	0/9	9/9	8/9
Standard and Poor 500 Index Negative Minute Return from May 2 2014 – Nov 13 2014					
$u_1 - u_9$	3/9	0/9	0/9	1/9	0/9
$u_{10} - u_{18}$	9/9	5/9	2/9	9/9	9/9

According to the results in table 4.1, exponential distribution is the least suitable one. In the range of threshold $u_{10} - u_{18}$, exponential distribution is rejected by Anderson-Darling test more frequently than other four distributions. In the range of threshold $u_1 - u_9$, exponential distribution is never accepted. Therefore, exponential distribution is not suitable distribution function in describing empirical log-returns.

However, Generalized Pareto distribution is most suitable distribution through Anderson-Darling test. Pareto distribution has higher frequencies of acceptance in the range of threshold $u_1 - u_9$; and it is never rejected in the range of threshold $u_{10} - u_{18}$. For MSCI World Index Negative Daily Return from July 3 1962 - Dec 31 2013, in the range $u_1 - u_9$, Pareto distribution is accepted quantic, while log-Weibull distribution is accepted twice and others are never accepted. Among the other range $u_{10} - u_{18}$, Pareto distribution is suitable in fitting data sample with 5% confidence threshold.

5. Application in Value at Risk and Violation Test

In this part, Pareto distribution, log-Weibull distribution, Power Law distribution, Weibull distribution, normal distribution and Student-t distribution are applied in value at risk.

5.1 What is Value at Risk and Violation Test

Value at Risk is widely used in financial risk management. It measures the risk of loss on a specific portfolio of financial assets. If a portfolio has a one-day 5% VaR of $1billion, then there is a 0.05 probability that the portfolio will fall in value by more than $ 1 billion over a one day period.

Normal test of value at risk measure is called violation tests. In the i-th period, compare the actual loss or gain of portfolio with the estimated value at risk level. If actual loss or gain of portfolio is less than negative value at risk level, then a violation occurs. $\begin{cases} I(i) = 0, -VaR < GL \\ I(i) = 1, -VaR \geq GL \end{cases}$, where GL is the actual gain or loss. Null hypothesis of the test is that value at risk is correctly estimated. $\sum_i^n I(i)$ is binomial distributed.

5.2 Application in Value at Risk

Rather than using statistic test, applying various distributions into sample data, estimating value at risk based on specified distribution and calculating frequency of violation would be better to test whether hypothesis distribution is suitable.

5.2.1 Method to calculate and test value at risk

We obtained the VaR of Pareto distribution, log-Weibull distribution, Power Law distribution and Weibull distribution by the same method.

Threshold percentages determine threshold returns. In general, 10% quantile threshold in 10-year historical returns is used to calculate one day 95% VaR and 5% quantile threshold in 10-year historical returns is used to calculate one day 99% VaR. $VaR_{100p} = F_{df}^{-1}\left(\frac{q}{p}\right)$, where p is quantile percentage, q is confidence interval for value at risk measure and F denotes cumulative distribution function.

The method to estimate value at risk by normal distribution is to obtain sample return based on data base years, model return distribution by using normal distribution and lose level, namely value at risk, which is obtained by inversing cumulative distribution functions. $VaR_{100p} = F_{df}^{-1}(p) * \sigma - u$, u is the mean of sample returns and σ is the standard deviation of sample returns.

The method to estimate value at risk by Student-t distribution is to obtain sample return based on formula that $VaR_{100p} = F_{df}^{-1}(p) * \sigma - u$, so $\begin{cases} VaR_{100p} = F_{df}^{-1}(p) * \sigma - u \\ VaR_{100q} = F_{df}^{-1}(q) * \sigma - u \end{cases}$, where p is quantile percentage and q is confidence level for value at risk measure.

In empirical world, mean is so small that can be neglected and degrees of freedom is 4, according to Fergusson& Platen (2005). Then we have, $VaR_{100q} = \frac{F_{df}^{-1}(q)}{F_{df}^{-1}(p)} * VaR_{100p}$.

5.2.2 Threshold Return and Time period

Indices including Dow Jones, S&P 500 index, MSCI world index, Russel 3000 index and Barclays Capital Bond Composite Global Index are tested. The portfolio consists of nine exchanged-traded funds--Energy Select Sector SPDR ETF, Financial Select Sector SPDR ETF, Utilities Select Sector SPDR ETF, Health Care Select Sector SPDR ETF, Industrial Select Sector SPDR ETF, Consumer Staples Select Sector SPDR ETF, Consumer Discrete Select SPDR ETF, Materials Select Sector SPDR ETF and Technology Select Sector SPDR ETF.

5.2.3 Violation Test

For this part, several indices are tested through violation test.

5.2.3.1 Violation Test for comparative long history

The test results are showed in table 5.2 and 5.3. According to the tables, 95% confidence interval estimated from normal distribution is not accepted in any one of seven tests and 99% confidence interval is accepted just one out of seven tests. Thus, normal distribution is not suitable in estimating value at risk even though it is recommended by Basel III.

Besides, the results of Pareto distribution, Power Law distribution, Log-Weibull distribution, Weibull distribution or Student t distribution are very similar. Half of null hypothesizes are based on the assumption that one specified distribution is accepted. Therefore, it is hard to determine which distribution is the most suitable in forecasting value at risk.

Table 5.2 Frequency of violations of VaR99 for long history data

Data	Time Period	Name	Distribution					
			Pareto	Power Law	Log-Weibull	Weibull	Normal	Student t
S&P500	Feb 7 1938-Dec 13 2014	Violation Percentage	1.1%	1.1%	1.1%	1.0%	1.8%	1.0%
		p-Value	22.2%	10.4%	7.1%	49.1%	0.0%	29.7%
ETF	Jan 2 2009-Dec 13 2014	Violation Percentage	0.6%	0.6%	0.6%	0.5%	1.6%	1.0%
		p-Value	7.6%	7.6%	7.6%	4.1%	1.6%	40.9%
Dow Jones	June14,1906-May 31, 2000	Violation Percentage	1.2%	1.2%	1.2%	1.1%	1.9%	1.1%
		p-Value	0.8%	0.3%	0.2%	15.5%	0.0%	1.8%
MSCI Index	Oct 3, 1981-Dec 13 2014	Violation Percentage	1.4%	1.4%	1.4%	1.3%	2.2%	1.2%
		p-Value	0.1%	0.0%	0.0%	0.9%	0.0%	3.2%
S&P 500	Nov 21,1988-Jul 31, 2003	Violation Percentage	1.5%	1.6%	1.5%	1.4%	1.8%	1.3%
		p-Value	0.8%	0.5%	0.8%	2.1%	0.0%	4.9%
Microsoft	Mar 4,1996-Jan 31, 2005	Violation Percentage	1.3%	1.3%	1.4%	1.1%	2.3%	1.2%
		p-Value	2.4%	1.7%	1.1%	18.3%	0.0%	11.2%
SPDR ETF	Jan 2 2009-Dec 13 2014	Violation Percentage	1.0%	1.0%	1.0%	0.8%	1.6%	1.4%
		p-Value	41.0%	41.0%	41.0%	28.4%	0.9%	4.6%

Table 5.3 Frequency of violations of VaR95 for long history data

Data	Time Period	Name	Distribution					
			Pareto	Power Law	Log-Weibull	Weibull	Normal	Student t
S&P500	Feb 7 1938- Dec 13 2014	Violation Percentage	5.1%	5.7%	5.2%	5.0%	4.5%	5.4%
		p-Value	21.8%	0.0%	6.3%	38.7%	0.1%	1.0%
ETF	Jan 2 2009- Dec 13 2014	Violation Percentage	4.5%	4.7%	4.7%	4.5%	3.9%	4.8%
		p-Value	19.0%	34.8%	30.4%	19.0%	2.2%	39.4%
Dow Jones	June14,1906- May 31, 2000	Violation Percentage	5.3%	5.9%	5.5%	5.3%	4.7%	5.7%
		p-Value	0.9%	0.0%	0.0%	3.8%	1.4%	0.0%
MSCI Index	Oct 3, 1981- Dec 13 2014	Violation Percentage	6.1%	6.7%	6.2%	5.9%	5.4%	6.2%
		p-Value	0.0%	0.0%	0.0%	0.0%	4.8%	0.0%
S&P 500	Nov 21,1988- Jul 31, 2003	Violation Percentage	5.8%	6.3%	5.7%	5.5%	4.9%	5.2%
		p-Value	4.1%	0.2%	6.0%	11.9%	43.8%	33.6%
Microsoft	Mar 4,1996- Jan 31, 2005	Violation Percentage	7.0%	7.6%	7.1%	6.8%	5.9%	7.6%
		p-Value	0.0%	0.0%	0.0%	0.0%	0.9%	0.0%
SPDR ETF	Jan 2 2009- Dec 13 2014	Violation Percentage	4.1%	4.6%	4.3%	4.0%	3.0%	5.3%
		p-Value	5.1%	26.1%	10.4%	3.9%	0.0%	25.1%

5.2.3.2 Violation Tests for period over financial crisis

These results are based on sample returns of 5 years including the financial crisis in 2008. None of the distributions is accepted based on the first three samples. Therefore these distributions are not good enough to estimate value at risk during economic recession.

Table 5.4 Frequency of violations of VaR99 during financial crisis period

Data	Time Period	Name	Distribution					
			Pareto	Power Law	Log-Weibull	Weibull	Normal	Student t
MSCI Index	Nov 4, 2005-Dec 31, 2010	Violation Percentage	2.53%	2.75%	2.75%	2.60%	3.72%	2.90%
		p-Value	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
S&P 500 Index	Jan 11, 2006-Dec 31, 2010	Violation Percentage	3.27%	3.43%	3.43%	3.11%	4.55%	3.67%
		p-Value	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Russel 3000 Index	Jan 11, 2006-Dec 31, 2010	Violation Percentage	3.43%	3.59%	3.59%	3.27%	4.78%	3.67%
		p-Value	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Barclays Index	Nov 8, 2005-Dec 31, 2010	Violation Percentage	1.33%	1.33%	1.33%	1.25%	1.48%	1.09%
		p-Value	9.83%	9.83%	9.83%	15.00%	3.70%	30.50%

Table 5.3 Frequency of violations of VaR95 during financial crisis period

Data	Time Period	Name	Distribution					
			Pareto	Power Law	Log-Weibull	Weibull	Normal	Student t
MSCI Index	Nov 4, 2005-Dec 31, 2010	Violation Percentage	8.8%	9.9%	9.1%	8.8%	8.4%	9.7%
		p-Value	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
S&P 500 Index	Jan 11, 2006-Dec 31, 2010	Violation Percentage	9.1%	10.2%	9.6%	9.0%	8.4%	9.6%
		p-Value	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Russel 3000 Index	Jan 11, 2006-Dec 31, 2010	Violation Percentage	9.4%	10.1%	9.6%	9.3%	8.7%	9.7%
		p-Value	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Barclays Index	Nov 8, 2005-Dec 31, 2010	Violation Percentage	5.7%	6.2%	5.7%	5.5%	4.7%	5.5%
		p-Value	11.4%	2.7%	11.4%	16.9%	33.0%	16.9%

5.2.3. Violation Tests conclusion

Even though Log Weibull distribution is statistically superior to Power Law distribution, but Power Law distribution has a similar quality in estimating value at risk empirically. Overall, Pareto distribution is the most suitable distribution in theoretical but it is indifferent from the other distributions when applied in reality

Appendices

Appendix A

Figure A-1 Returns Distribution of Down Jones

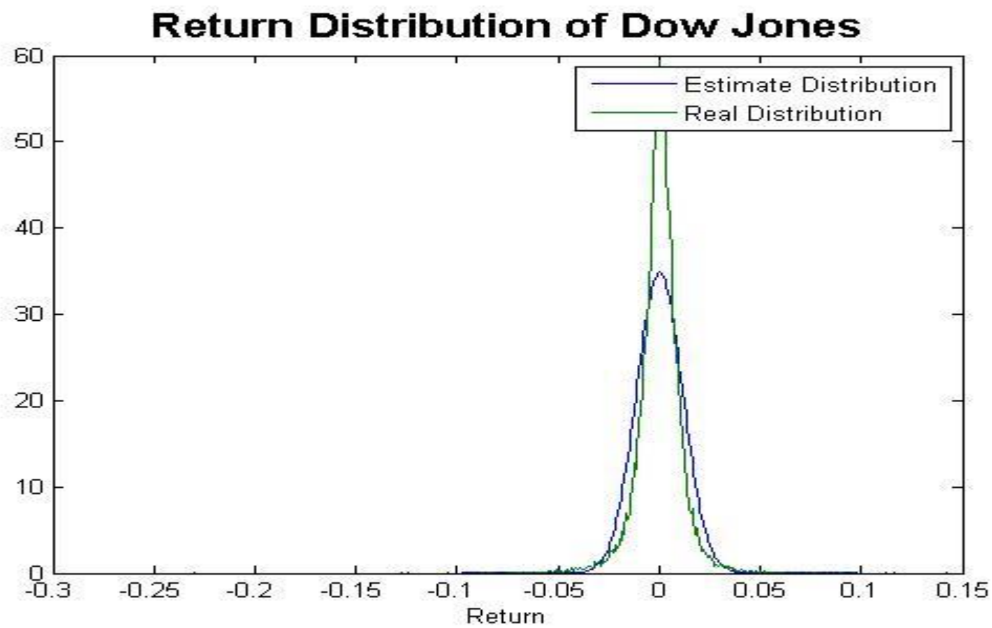


Figure A-2 Daily Return Distribution of S&P 500

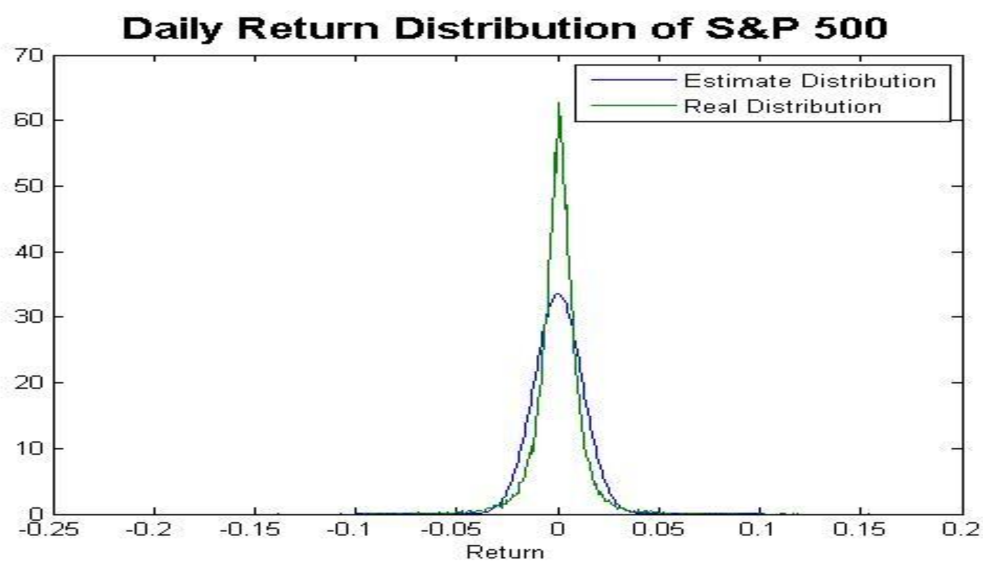


Figure A-3 Return Distribution of MSCI Index

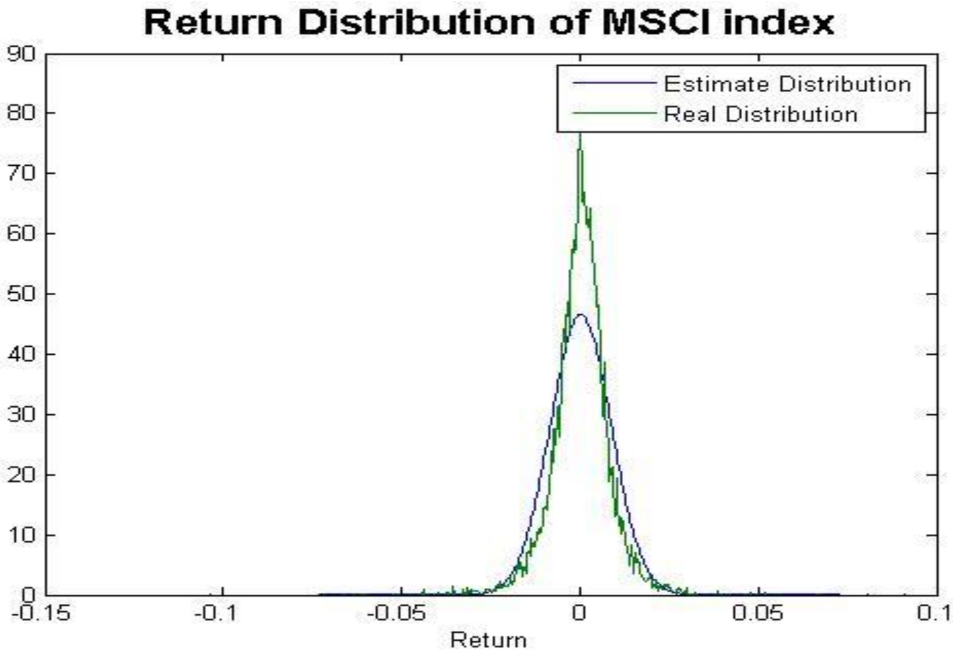
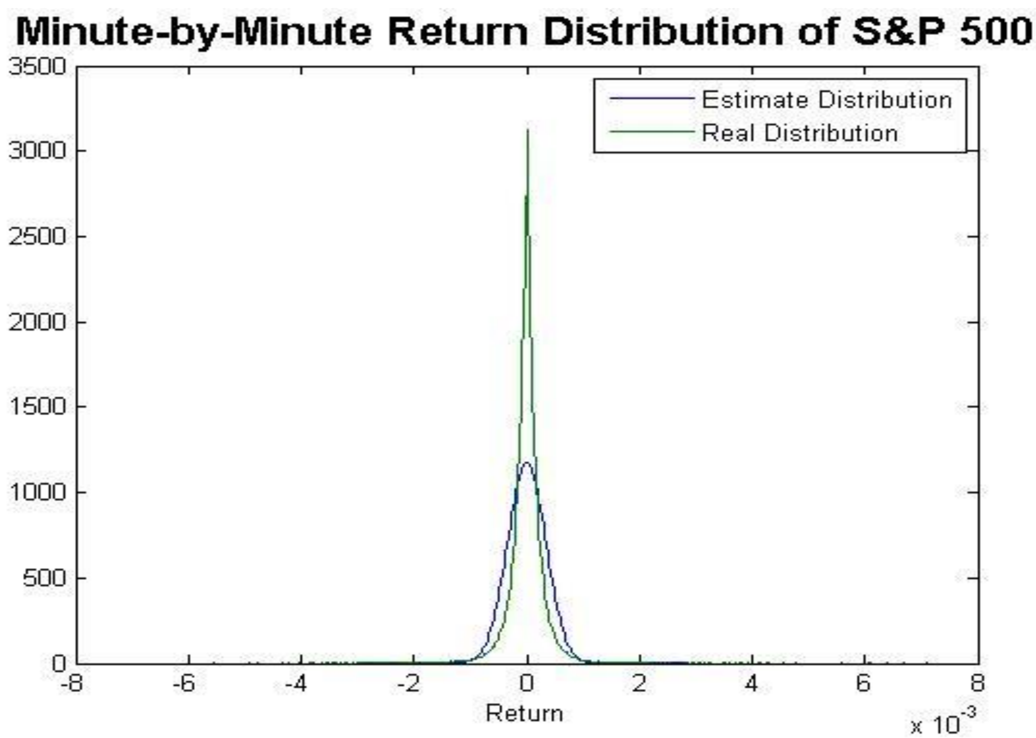


Figure A-4 Minute-by-Minute Return Distribution of S&P 500



Appendix B

Table B-1 Probability levels, corresponding lower thresholds and size of sub-sample beyond the thresholds of daily S&P 500 Index from Jan 02 1926 to Nov 13 2013

S&P 500 (Daily)				
	Positive Return		Negative Return	
Threshold (probability)	Number of Data used	Threshold (Min. Return *100)	Number of Data used	Threshold (Min. Return *100)
0	6849	0.0006	6068	0.0008
0.1	6165	0.0843	5461	0.0839
0.2	5480	0.1744	4854	0.1691
0.3	4796	0.2617	4249	0.2562
0.4	4110	0.3678	3641	0.3571
0.5	3425	0.4780	3034	0.4864
0.6	2740	0.6160	2427	0.6380
0.7	2055	0.7912	1820	0.8275
0.8	1370	1.0427	1214	1.0828
0.9	685	1.4989	607	1.5504
0.925	514	1.6882	455	1.7330
0.95	342	1.9704	303	2.0201
0.96	274	2.1589	243	2.2235
0.97	205	2.3418	182	2.4181
0.98	137	2.7031	121	2.7311
0.99	68	3.4745	61	3.4106
0.9925	51	3.8416	46	3.8340
0.995	34	4.1059	30	4.4223

Figure B-1 Complementary Cumulative Distribution Function of daily S&P 500Index

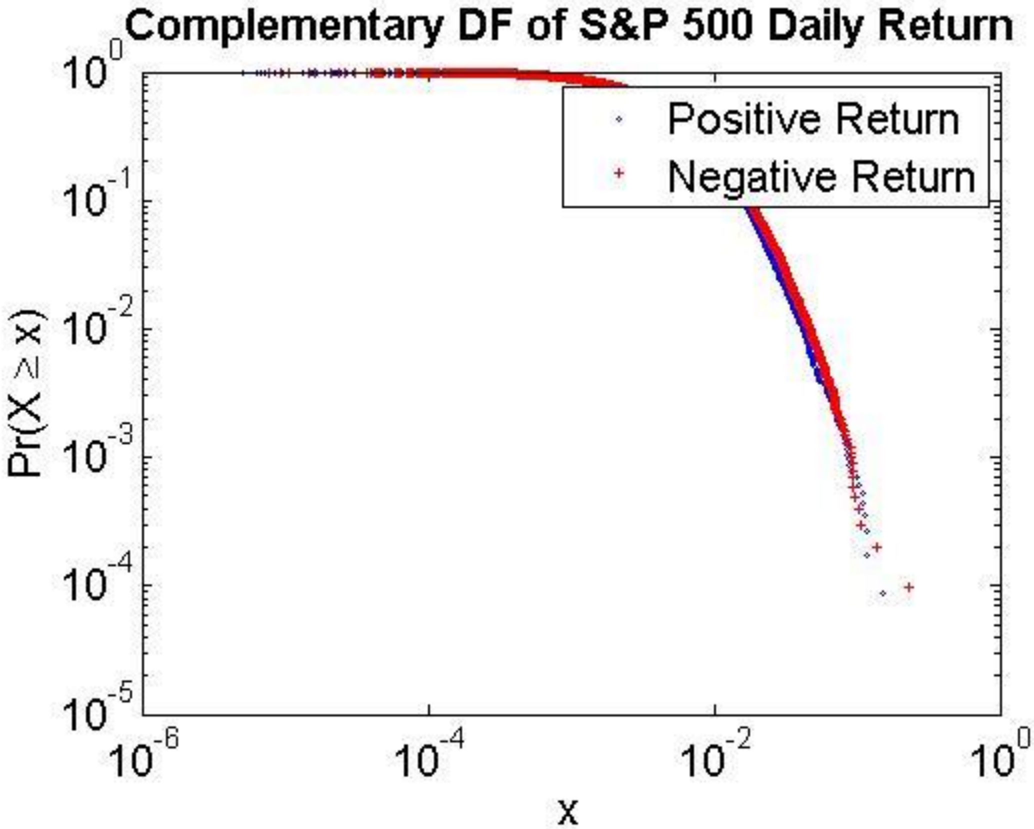


Table B-2 Probability levels, corresponding lower thresholds and size of sub-sample beyond the thresholds of MSCI Index from July 3 1962 to Nov 13 2014

MSCI				
	Positive Return		Negative Return	
Threshold (probability)	Number of Data used	Threshold (Min. Return *100)	Number of Data used	Threshold (Min. Return *100)
0	5975	0.0006	5236	0.0011
0.1	5378	0.0814	4712	0.0727
0.2	4780	0.1608	4189	0.1506
0.3	4183	0.2469	3665	0.2296
0.4	3585	0.3286	3142	0.3231
0.5	2988	0.4231	2618	0.4286
0.6	2390	0.5340	2094	0.5409
0.7	1793	0.6755	1571	0.6949
0.8	1195	0.8767	1047	0.9066
0.9	598	1.2221	524	1.3172
0.925	448	1.3774	393	1.4655
0.95	299	1.6079	262	1.7313
0.96	239	1.7247	209	1.8637
0.97	179	1.9258	157	2.0897
0.98	120	2.2262	105	2.4274
0.99	60	2.7042	52	3.1551
0.9925	45	2.8596	39	3.5340
0.995	30	3.1433	26	4.1313

Figure B-2 Complementary Cumulative Distribution Function of MSCI Index

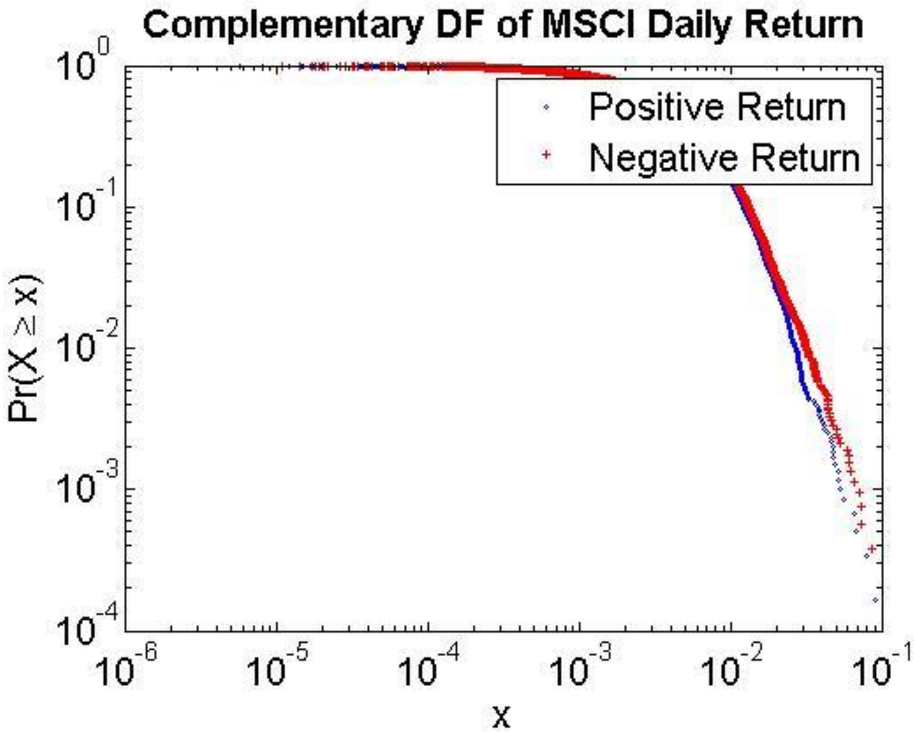
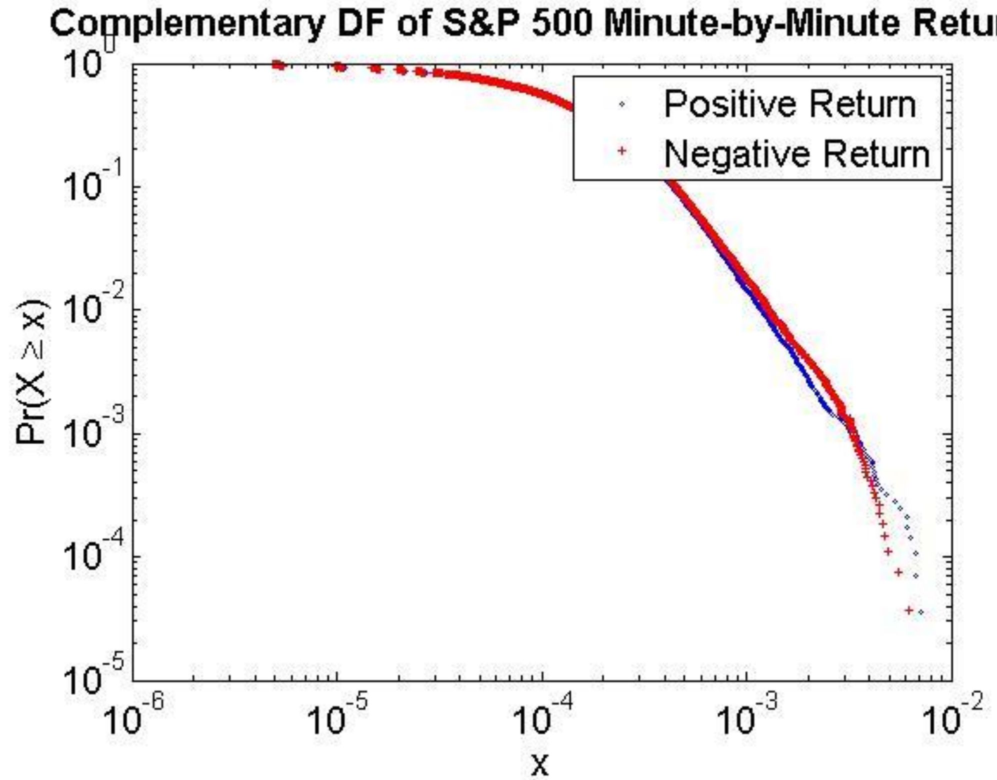


Table B-3 Probability levels, corresponding lower thresholds and size of sub-sample beyond the thresholds of minute by minute MSCI Index from May 2 2014to Nov 13 2014

S&P500 (Minutely)				
	Positive Return		Negative Return	
Threshold (probability)	Number of Data used	Threshold (Min. Return *100)	Number of Data used	Threshold (Min. Return *100)
0	28359	0.0005	26794	0.0005
0.1	25523	0.0020	24115	0.0020
0.2	22687	0.0036	21435	0.0039
0.3	19851	0.0060	18756	0.0061
0.4	17015	0.0086	16076	0.0089
0.5	14180	0.0116	13397	0.0120
0.6	11344	0.0154	10718	0.0159
0.7	8508	0.0207	8038	0.0214
0.8	5672	0.0283	5359	0.0294
0.9	2836	0.0435	2679	0.0451
0.925	2127	0.0500	2010	0.0519
0.95	1418	0.0602	1340	0.0626
0.96	1134	0.0666	1072	0.0695
0.97	851	0.0744	804	0.0791
0.98	567	0.0884	536	0.0948
0.99	284	0.1183	268	0.1293
0.9925	213	0.1334	201	0.1481
0.995	142	0.1604	134	0.1781

Figure B-3 Complementary Cumulative Distribution Function of minute by minute S&P 500 Index



Appendix C

Table C-1 Dow Jones positive returns distribution

	Dow Jones positive returns distribution									
Threshold (probability)	Pareto Distribution		Weibull Distribution		exponential distribution		Log-Weibull distribution		Power-Law distribution	
	AD distance	p-Value	AD distance	p-Value	AD distance	p-Value	AD distance	p-Value	AD distance	p-Value
0	55.33	0.00	19.93	0.00	41.35	0.00	52.17	0.00	3795.19	0.00
0.1	27.76	0.00	15.61	0.00	16.09	0.00	92.37	0.00	1385.67	0.00
0.2	21.85	0.00	16.45	0.00	13.49	0.00	62.02	0.00	842.58	0.00
0.3	12.64	0.00	13.95	0.00	12.40	0.00	38.83	0.00	492.95	0.00
0.4	7.63	0.00	12.60	0.00	15.87	0.00	24.20	0.00	290.24	0.00
0.5	5.51	0.00	12.34	0.00	19.96	0.00	13.36	0.00	171.24	0.00
0.6	5.25	0.00	13.15	0.00	23.20	0.00	4.24	0.01	99.36	0.00
0.7	0.90	**0.41**	8.29	0.00	33.03	0.00	1.56	**0.16**	31.19	0.00
0.8	0.30	**0.94**	4.34	0.01	25.63	0.00	1.73	**0.13**	11.90	0.00
0.9	0.29	**0.95**	2.27	**0.07**	19.90	0.00	0.43	**0.82**	1.12	**0.30**
0.925	0.26	**0.96**	1.83	**0.11**	11.67	0.00	0.39	**0.86**	1.58	**0.16**
0.95	0.25	**0.97**	1.20	**0.27**	6.11	0.00	0.35	**0.90**	1.55	**0.16**
0.96	0.28	**0.95**	0.92	**0.40**	5.21	0.00	0.40	**0.85**	1.06	**0.33**
0.97	0.55	**0.69**	0.77	**0.50**	3.81	0.01	0.85	**0.45**	1.20	**0.27**
0.98	0.22	**0.98**	0.47	**0.78**	4.07	0.01	0.19	**0.99**	0.19	**0.99**
0.99	0.26	**0.96**	0.38	**0.87**	1.40	**0.20**	0.28	**0.95**	0.35	**0.89**
0.9925	0.38	**0.87**	0.61	**0.64**	1.10	**0.31**	0.31	**0.93**	0.49	**0.76**
0.995	0.46	**0.78**	0.28	**0.95**	2.11	**0.08**	0.29	**0.95**	0.57	**0.67**

Table C-2 Dow Jones negative returns distribution

	Dow Jones negative return distribution									
Threshold (probability)	Pareto Distribution		Weibull Distribution		exponential distribution		Log-Weibull distribution		Power-Law distribution	
	AD distance	p-Value	AD distance	p-Value	AD distance	p-Value	AD distance	p-Value	AD distance	p-Value
0	18.79	0.00	20.27	0.00	18.13	0.00	30.98	0.00	3660.90	0.00
0.1	9.83	0.00	16.37	0.00	20.02	0.00	74.76	0.00	1230.95	0.00
0.2	4.74	0.00	12.93	0.00	26.20	0.00	57.54	0.00	673.88	0.00
0.3	4.88	0.00	13.54	0.00	26.78	0.00	36.15	0.00	438.27	0.00
0.4	3.71	0.01	12.98	0.00	30.31	0.00	19.90	0.00	267.57	0.00
0.5	3.49	0.02	13.44	0.00	31.77	0.00	8.58	0.00	162.74	0.00
0.6	0.57	**0.67**	8.62	0.00	39.99	0.00	5.59	0.00	69.46	0.00
0.7	0.44	**0.81**	6.66	0.00	33.03	0.00	2.50	**0.05**	35.43	0.00
0.8	0.44	**0.81**	3.03	0.03	28.52	0.00	1.99	**0.09**	10.53	0.00
0.9	0.23	**0.98**	1.69	**0.14**	8.20	0.00	0.80	**0.48**	6.60	0.00
0.925	0.31	**0.93**	1.58	**0.16**	5.81	0.00	0.48	**0.76**	4.70	0.00
0.95	0.62	**0.63**	0.58	**0.67**	5.81	0.00	1.23	**0.26**	1.78	**0.12**
0.96	0.37	**0.88**	1.07	**0.32**	2.57	**0.05**	0.39	**0.86**	2.60	0.04
0.97	0.36	**0.89**	0.69	**0.56**	3.61	0.01	0.54	**0.70**	0.87	**0.43**
0.98	0.62	**0.63**	1.24	**0.25**	1.97	**0.10**	0.19	**0.99**	1.41	**0.20**
0.99	0.61	**0.64**	0.83	**0.46**	3.08	0.02	0.65	**0.60**	0.61	**0.64**
0.9925	0.29	**0.94**	0.71	**0.55**	3.01	0.03	0.31	**0.93**	0.35	**0.90**
0.995	0.32	**0.93**	0.64	**0.61**	2.45	**0.05**	0.33	**0.91**	0.36	**0.88**

Table C-3 MSCI positive returns distribution

	MSCI Positive Return Distribution									
Threshold (probability)	Pareto Distribution		Weibull Distribution		exponential distribution		Log-Weibull distribution		Power-Law distribution	
	AD distance	p-Value	AD distance	p-Value	AD distance	p-Value	AD distance	p-Value	AD distance	p-Value
0	27.61	0.00	0.00	0.00	26.38	0.00	22.90	0.00	1895.01	0.00
0.1	16.02	0.00	0.00	0.00	12.28	0.00	41.51	0.00	648.51	0.00
0.2	10.26	0.00	0.00	0.00	6.57	0.00	22.35	0.00	377.91	0.00
0.3	3.54	0.01	0.01	0.00	2.85	0.03	16.36	0.00	206.00	0.00
0.4	2.67	0.04	0.02	0.00	3.50	0.02	9.00	0.00	130.71	0.00
0.5	1.41	**0.20**	0.03	0.00	4.49	0.01	5.65	0.00	75.96	0.00
0.6	0.70	**0.56**	0.07	0.00	5.34	0.00	3.81	0.01	41.20	0.00
0.7	0.55	**0.69**	0.11	0.00	6.47	0.00	3.85	0.01	19.20	0.00
0.8	0.36	**0.89**	0.12	0.00	5.62	0.00	0.62	**0.63**	8.07	0.00
0.9	0.29	**0.94**	0.20	**0.08**	4.08	0.01	0.70	**0.56**	2.15	**0.08**
0.925	0.56	**0.68**	0.20	**0.06**	3.31	0.02	0.33	**0.91**	1.80	**0.12**
0.95	0.47	**0.78**	0.58	**0.39**	3.29	0.02	0.61	**0.64**	0.72	**0.54**
0.96	0.62	**0.63**	0.29	**0.87**	1.88	**0.11**	0.34	**0.90**	1.37	**0.21**
0.97	0.55	**0.70**	0.33	**0.57**	2.32	**0.06**	0.47	**0.78**	0.69	**0.57**
0.98	0.48	**0.76**	0.11	**0.40**	3.65	0.01	0.80	**0.48**	0.61	**0.64**
0.99	0.86	**0.44**	0.63	**0.96**	3.27	0.02	0.69	**0.57**	1.33	**0.22**
0.9925	1.17	**0.28**	0.82	**0.87**	1.75	**0.13**	0.69	**0.57**	1.02	**0.35**
0.995	0.32	**0.92**	0.90	**0.97**	0.37	**0.88**	0.49	**0.75**	0.92	**0.40**

Table C-4 MSCI negative returns distribution

	MSCI Negative Return Distribution									
Threshold (probability)	Pareto Distribution		Weibull Distribution		exponential distribution		Log-Weibull distribution		Power-Law distribution	
	AD distance	p-Value	AD distance	p-Value	AD distance	p-Value	AD distance	p-Value	AD distance	p-Value
0	11.51	0.00	6.13	0.00	6.67	0.00	21.11	0.00	1544.58	0.00
0.1	8.24	0.00	5.75	0.00	4.74	0.00	36.81	0.00	560.66	0.00
0.2	4.98	0.00	4.88	0.00	4.12	0.01	25.37	0.00	315.62	0.00
0.3	4.23	0.01	5.19	0.00	4.96	0.00	15.00	0.00	200.65	0.00
0.4	2.16	**0.08**	4.45	0.01	6.75	0.00	9.77	0.00	113.56	0.00
0.5	0.97	**0.37**	3.04	0.03	10.15	0.00	8.55	0.00	57.38	0.00
0.6	1.42	**0.20**	4.03	0.01	8.48	0.00	3.49	0.02	38.96	0.00
0.7	0.96	**0.38**	3.45	0.02	9.26	0.00	1.30	**0.23**	18.28	0.00
0.8	1.03	**0.34**	3.19	0.02	7.42	0.00	0.80	**0.48**	9.16	0.00
0.9	0.28	**0.95**	1.37	**0.21**	10.15	0.00	0.28	**0.95**	0.36	**0.89**
0.925	0.60	**0.65**	1.70	**0.14**	6.35	0.00	0.50	**0.75**	0.91	**0.41**
0.95	0.94	**0.39**	0.21	**0.99**	6.61	0.00	0.58	**0.67**	0.82	**0.46**
0.96	0.43	**0.82**	0.20	**0.99**	3.01	0.03	0.42	**0.83**	0.44	**0.81**
0.97	0.32	**0.93**	0.30	**0.93**	1.67	**0.14**	0.39	**0.86**	0.57	**0.68**
0.98	0.26	**0.97**	0.37	**0.88**	0.51	**0.73**	0.19	**0.99**	0.86	**0.44**
0.99	0.29	**0.94**	0.30	**0.94**	0.38	**0.87**	0.30	**0.94**	0.48	**0.77**
0.9925	0.54	**0.70**	0.34	**0.90**	0.65	**0.60**	0.51	**0.73**	0.52	**0.73**
0.995	0.57	**0.68**	0.57	**0.68**	0.61	**0.64**	0.59	**0.65**	0.67	**0.58**

Table C-5 Minute by minute S&P 500 positive returns distribution

	S&P 500 Minute by minute Positive Distribution									
Threshold (probability)	Pareto Distribution		Weibull Distribution		exponential distribution		Log-Weibull distribution		Power-Law distribution	
	AD distance	p-Value	AD distance	p-Value	AD distance	p-Value	AD distance	p-Value	AD distance	p-Value
0	27.61	0.00	7.46	0.00	26.38	0.00	22.90	0.00	1895.01	0.00
0.1	16.02	0.00	6.33	0.00	12.28	0.00	41.51	0.00	648.51	0.00
0.2	10.26	0.00	5.23	0.00	6.57	0.00	22.35	0.00	377.91	0.00
0.3	3.54	0.01	3.61	0.01	2.85	0.03	16.36	0.00	206.00	0.00
0.4	2.67	0.04	3.40	0.02	3.50	0.02	9.00	0.00	130.71	0.00
0.5	1.41	**0.20**	3.08	0.03	4.49	0.01	5.65	0.00	75.96	0.00
0.6	0.70	**0.56**	2.17	**0.07**	5.34	0.00	3.81	0.01	41.20	0.00
0.7	0.55	**0.69**	1.83	**0.11**	6.47	0.00	3.85	0.01	19.20	0.00
0.8	0.36	**0.89**	1.77	**0.12**	5.62	0.00	0.62	**0.63**	8.07	0.00
0.9	0.29	**0.94**	1.42	**0.20**	4.08	0.01	0.70	**0.56**	2.15	**0.08**
0.925	0.56	**0.68**	1.42	**0.20**	3.31	0.02	0.33	**0.91**	1.80	**0.12**
0.95	0.47	**0.78**	0.67	**0.58**	3.29	0.02	0.61	**0.64**	0.72	**0.54**
0.96	0.62	**0.63**	1.15	**0.29**	1.88	**0.11**	0.34	**0.90**	1.37	**0.21**
0.97	0.55	**0.70**	1.06	**0.33**	2.32	**0.06**	0.47	**0.78**	0.69	**0.57**
0.98	0.48	**0.76**	1.87	**0.11**	3.65	0.01	0.80	**0.48**	0.61	**0.64**
0.99	0.86	**0.44**	0.62	**0.63**	3.27	0.02	0.69	**0.57**	1.33	**0.22**
0.9925	1.17	**0.28**	0.43	**0.82**	1.75	**0.13**	0.69	**0.57**	1.02	**0.35**
0.995	0.32	**0.92**	0.34	**0.90**	0.37	**0.88**	0.49	**0.75**	0.92	**0.40**

Table C-6 Minute by minute S&P 500 negative returns distribution

	S&P 500 Minute by minute Negative Distribution									
Threshold (probability)	Pareto Distribution		Weibull Distribution		exponential distribution		Log-Weibull distribution		Power-Law distribution	
	AD distance	p-Value	AD distance	p-Value	AD distance	p-Value	AD distance	p-Value	AD distance	p-Value
0	68.50	0.00	87.21	0.00	337.38	0.00	668.70	0.00	4167.17	0.00
0.1	16.51	0.00	54.47	0.00	216.51	0.00	262.04	0.00	2368.40	0.00
0.2	6.74	0.00	46.39	0.00	195.16	0.00	131.62	0.00	1417.39	0.00
0.3	6.69	0.00	49.90	0.00	175.54	0.00	58.35	0.00	921.26	0.00
0.4	2.71	0.04	38.32	0.00	183.82	0.00	42.79	0.00	513.19	0.00
0.5	2.14	**0.08**	31.61	0.00	170.37	0.00	32.48	0.00	296.48	0.00
0.6	2.73	0.04	31.26	0.00	142.03	0.00	16.00	0.00	178.76	0.00
0.7	1.74	**0.13**	26.41	0.00	130.23	0.00	6.07	0.00	79.56	0.00
0.8	1.16	**0.28**	19.61	0.00	106.91	0.00	2.23	**0.07**	26.88	0.00
0.9	0.28	**0.95**	10.37	0.00	78.32	0.00	0.23	**0.98**	1.54	**0.17**
0.925	0.34	**0.90**	7.32	0.00	59.11	0.00	0.28	**0.95**	0.77	**0.51**
0.95	0.44	**0.81**	3.69	0.01	35.95	0.00	0.40	**0.85**	0.62	**0.63**
0.96	0.51	**0.74**	2.99	0.03	27.20	0.00	0.45	**0.79**	0.57	**0.68**
0.97	0.60	**0.65**	2.36	**0.06**	17.12	0.00	0.49	**0.76**	0.84	**0.46**
0.98	0.80	**0.48**	1.37	**0.21**	8.83	0.00	0.63	**0.62**	1.03	**0.34**
0.99	1.70	**0.14**	0.41	**0.84**	2.96	0.03	1.44	**0.19**	1.50	**0.18**
0.9925	1.25	**0.25**	0.54	**0.71**	1.34	**0.22**	1.43	**0.19**	1.63	**0.15**
0.995	0.17	**1.00**	0.25	**0.97**	0.27	**0.96**	0.75	**0.52**	2.44	**0.05**

References

1. Mandelbrot, Benoit(1963), “The variation of certain speculative prices”, Journal of Business, 1963, 36, pp.392–417.
2. Fama E.F(1965), “The behavior of stock market prices”. The Journal of Business, 1965, 3 8, 34–105.
3. Fama E.F(1965), “Mandelbrot and the Stable Paretian Hypothesis”, The Journal of Business, Vol. 36, No. 4. (Oct., 1963), pp. 420-429.
4. 4.Stuart, A. and Ord, K,(1994), “Kendall’s Advances Theory of Statistics”,1994 (Wiley: New York).
5. Sornette, D., Simonetti, P. and Andersen, J.V.(2000), “q-Field theory for portfolio optimization: ‘fat tails’ and non-linear correlations”. Physics Report, 2000, 335, 19–92.
6. Paolella, Marc S and Mittnik, Stefan (1997), “Econometric modeling in the presence of heavy-tailed innovations: a survey of some recent advances”, Journal of banking & finance, Volume 31, Issue 11, pp. 3462 – 3485
7. Anderson, T.W. and Darling (1952), “D.A., Asymptotic theory of certain ‘goodness of fit’ criteria.” Annals of Mathematical Statistics, 1952, 23, 193–212.
8. Kevin Fergusson & Eckhard Platen (2006), “On the Distributional Characterization of Daily Log-Returns of a World Stock Index”, Applied Mathematical Finance Volume 13, Issue 1, 2006, pp. 19-38
9. Y. Malevergne , V. Pisarenko & D. Sornette (2005), “Empirical distributions of stock returns: between the stretched exponential and the power law?”, Quantitative Finance, Quantitative Finance Volume 5, Issue 4, 2005, pages 379-401